

**\* NOTICES \***

JPO and NCIP are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

**CLAIMS**

---

[Utility model registration claim]

[Claim 1] In the voice recognition unit which will output the talk back signal corresponding to this sound signal, and will answer with voice if the signal of the voice which the speaker uttered inputs The language judging section used which judges the language from the sound signal inputted as the voice input section (1) at least (2), The speech recognition section (3) which distinguishes the contents of the inputted sound signal; and the control section which controls a device according to a speech recognition result (4), It has the reply signal generating section for an audio response (5). The speech recognition section (3) Two or more languages can be recognized from the inputted sound signal. Moreover, the reply signal generating section (5) The audio response by two or more languages has become possible, and the judgment signal outputted from the language judging section (2) used inputs into the speech recognition section (3). The voice recognition unit characterized by for the judgment signal which recognition object language was selected and was outputted from the language judging section (2) used inputting into the reply signal generating section (5), and being constituted so that audio response language may be selected.

[Claim 2] In the voice recognition unit which will output the talk back signal corresponding to this sound signal, and will answer with voice if the signal of the voice which the speaker uttered inputs The speech recognition section which distinguishes the contents of the sound signal inputted as the voice input section (1) at least (3), It has the control section (4) which controls a device according to a speech recognition result, the reply signal generating section for an audio response (5), and the setting section (6) of the language used. The speech recognition section (3) Two or more languages can be recognized from the inputted sound signal. Moreover, the reply signal generating section (5) The audio response by two or more languages has become possible. Either [ at least ] said speech recognition section (3) or said reply signal generating section (5) The voice recognition unit characterized by being constituted according to the selection signal outputted from the language setting section (6) used so that recognition object language or audio response language may be chosen.

---

[Translation done.]

## DETAILED DESCRIPTION

---

[Detailed explanation of a design] [0001]

[Industrial Application]

This design is related with the voice recognition unit which answers with voice to utterance of a speaker. While operators, such as an automobile, operate, when devices, such as an air conditioner, audio equipment, car navigation, a land mobile radiotelephone, and auto-cruise, are operated manually, trouble is caused to a safety operation and there is a possibility of causing accident. Then, it is desirable to direct actuation of these devices with voice, and to answer with voice to the directed actuation at a device side, and for a speaker to enable it to check because of the improvement in operability.

[0002]

[Description of the Prior Art]

Not only the object for automobiles but a voice recognition unit operates a device according to the contents which have recognized a speaker's voice and have been recognized further. And a device side to tone is emitted to a speaker, and a result is answered.

[0003]

Drawing 5 is a flow chart which shows the response actuation in the conventional voice recognition unit. If it is the case where audio equipment is operated, and it will utter with a "play", voice input will be carried out in step S1 and playback actuation will be directed first, in step S2, a sound signal will be inputted through a microphone, and pretreatment will be performed so that it may be suitable for speech recognition at step S3.

[0004]

Subsequently, as shown in step S4, in the speech recognition section, speech recognition is performed and the judgment of a recognition result is performed in step S5. When the word "play" has been recognized, after answering with an audible tone etc. and telling a speaker about what has been recognized in step S6, play actuation is started in step S7.

[0005]

On the other hand, when a speaker's words cannot be recognized, in step S8, the tone which tells that is outputted and a speaker is urged to carry out voice input again.

[0006]

[Problem(s) to be Solved by the Device]

However, in the response by such an audible tone etc., only those about two kinds which were able to carry out speech recognition or are not made can be expressed. Since in the case of an automobile etc. many devices are equipped and it is necessary to operate much switches, the contents of the actuation performed by switch actuation are also various. Therefore, it is desirable for a response to be also able to answer with voice concretely according to the contents of the operator guidance instead of an audible tone.

[0007]

For example, for example like "playing a cassette tape", if the directions voice is recognized and a play is moreover possible when a speaker utters with a "play", when speech recognition cannot be answered and carried out with voice, if an audio response (talk back) can be carried out, device actuation can be more correctly performed to a dialogue and operability will improve like "talk once again."

[0008]

Moreover, about conditions, such as temperature and airflow, if a speaker can be told about conditions, such as the present rate, by the audio response in the case of auto-cruise, more [ in the case of an air conditioner ] will become simple [ actuation ].

[0009]

By the way, although speech recognition is performed in Japanese, it usually needs to make possible speech recognition of the case of an export vehicle in a partner language. Moreover, also when answering with voice like this design, it is necessary to answer in a partner language about the car for export.

[0010]

However, in order to enable it to perform speech recognition and the talk back in partner languages, such as English, only about an export vehicle, it is necessary to make speech recognition of English possible and to develop independently the recognition equipment which can moreover also perform the talk back in English, and there is a problem of becoming cost quantity.

[0011]

On the other hand, when internationalization progressed and a rental car etc. is assumed, if Japanese people may operate the same car, a foreigner may operate. in such a case — Japanese — English — speech recognition — if the talk back cannot be carried out, the desired end cannot be attained but it is inconvenient.

[0012]

The technical technical problem of this design is to improve user-friendliness paying attention to such a problem by making speech recognition and an audio response possible by two or more languages beforehand, and realize the cheap high equipment of versatility.

[0013]

[Means for Solving the Problem]

Drawing 1 is a block diagram explaining the basic configuration of the voice recognition unit by this design. The voice recognition unit of claim 1 is this sound signal, when the signal of the voice which the speaker uttered inputs.

It is alike, a corresponding talk back signal is outputted, and it has the voice input section 1, the language judging section 2 used which judges the language from the inputted sound signal, the speech recognition section 3 which distinguishes the contents of the inputted sound signal, the control section 4 which controls a device according to a speech recognition result, and the reply signal generating section 5 for an audio response at least like drawing 1 (1) in the voice recognition unit which answers with voice.

[0014]

The aforementioned speech recognition section 3 can recognize two or more languages from the inputted sound signal, and the reply signal generating section 5 can carry out the audio response of it in two or more languages. And the judgment signal which the judgment signal outputted from the language judging section 2 used inputted into the speech recognition section 3, and recognition object language was selected, and was outputted from the language judging section 2 used inputs into the reply signal generating section 5, and it is constituted so that audio response language may be selected.

[0015]

The talk back signal corresponding to this sound signal is outputted, it sets to the voice recognition unit which answers with voice, and claim 2 is drawing 1 (2), when the signal of the voice which the speaker uttered inputs.

\*\* -- it has the voice input section 1, the speech recognition section 3 which distinguishes the contents of the inputted sound signal, the control section 4 which controls a device according to a speech recognition result, the reply signal generating section 5 for an audio response, and the setting section 6 of the language used at least like.

[0016]

The speech recognition section 3 can recognize two or more languages from the inputted sound signal, and is a reply signal.

The audio response by two or more languages is possible for the generating section 5. And according to the selection signal outputted from the language setting section 6 used, either [ at least ] said speech recognition section 3 or said reply signal generating section 5 is constituted so that recognition object language or audio response language may be chosen.

[0017]

[Function]

The speech recognition section 3 of this design has the speech recognition function for every language, and has the composition that the speech recognition function of the language judged by the judgment signal from the language judging section 2 used operates. Therefore, two or more languages can be recognized from the inputted sound signal.

[0018]

Moreover, since the reply signal generating section 5 has the audio response function according to operating instructions for every language, If the language is automatically judged from the inputted sound signal in the language judging section 2 used, a judgment signal inputs into the speech recognition section 3 and the reply signal generating section 5 and the language used is specified In the speech recognition section 3, speech recognition only of the specified language is chosen and carried out, in the reply signal generating section 5, the specified language is chosen and an audio response is performed.

[0019]

Thus, since voice only inputs from the voice input section 1, the language is judged and speech recognition and an audio response are performed in the language, speech recognition and when it can carry out an audio response and the equipment for export and the race of two or more nations use it, it is convenient at two or more languages. Moreover, it is not necessary to consider as equipment with the export place only for languages, and since versatility is high, it is cheaply realizable.

[0020]

In addition, the first sound signal inputted by the initial state in the language judging section 2 used

Although it comes out, the classification of language is judged and there are technique to which only the language is confirmed henceforth, and the technique of judging the classification of the language whenever a sound signal inputs, and inputting a judgment signal into the speech recognition section 3 and the reply signal generating section 5, and specifying the language used, this design shall contain both.

[0021]

The speech recognition section 3 in claim 2 can also recognize two or more languages from the inputted sound signal. Moreover, the reply signal generating section 5 The audio response has become possible by two or more languages, and the setting signal outputted from the language setting section 6 used inputs into the speech recognition section 3. The setting signal which recognition object language was chosen or was outputted from the language setting section 6 used inputs into the reply signal generating section 5, and it has the composition that audio response language is chosen.

[0022]

Therefore, since it is not necessary to operate it unless the language used will be changed, once it does so the same effectiveness as the case of claim 1 since it can respond to two or more languages, although the actuation which chooses the language used beforehand artificially is required, and also it sets up the language used and constituting becomes simple in favor of little [ the frequency of modification of the language used ] equipment, it can realize more cheaply.

[0023]

[Example] Next, an example explains how the voice recognition unit by this design is materialized in practice. Drawing 2 is the block diagram showing the example which carried out the voice recognition unit of claim 1 in the automobile. Ahead of the front seat 8 of an automobile 7, Microphones 1a and 1b are arranged in the both sides of a drivers side and a passenger side as a voice input means.

[0024]

And the sound signal detected with Microphones 1a and 1b is inputted into the judgment section 2 of the language used through the pretreatment circuit 9 which performs noise rejection etc., and the classification of language in a front seat 8 about which the speaker spoke is distinguished. In the language judging section 2 used, if a speaker's language used is distinguished, the distinction signal will be inputted into the speech recognition section 3 and the reply signal generating section 5, and only the judged language will become effective.

[0025]

Audio equipment A1, an air conditioner A2, land mobile radiotelephone A3, navigation equipment A4, Auto-cruise An, etc. are carried in this automobile as equipment.

[0026]

This example shall judge the classification of language with the first sound signal inputted by the initial state, and it shall have henceforth composition which confirms only that language. If the speaker of a front seat 8 is the case where a device is operated in Japanese, will start an engine and will utter with "NIHONGO" first in an initial state now, the sound signal inputted from Microphones 1a or 1b will be inputted into the language judging section 2 used through the pretreatment circuit 9, and it will judge that the classification of the language used is "Japanese."

[0027]

And only Japanese is confirmed by this judgment signal in the speech recognition section 3 and the reply signal generating section 5, and comparison collating of the Japanese word thru/or text registered into the speech recognition section 3, and the inputted word thru/or text is carried out by it

[0028]

For example, if it is the case where the cassette tape of audio equipment A1 is played and a speaker will utter with a "tape play" first, through the pretreatment circuit 9 and the language judging section 2 used, the sound signal inputted from Microphones 1a or 1b will be inputted into the speech recognition section 3, and comparison collating will be carried out with the dictionary of Japanese registered into the speech recognition section 3.

[0029]

Although the dictionary of other languages, such as English, is also registered, the table on which other languages are registered is not chosen, but only the table on which Japanese is registered is chosen by the judgment signal inputted from the language judging section 2 used, and comparison collating of it is carried out with the vocabulary registered beforehand.

[0030]

Consequently, it is recognized that it is the playback instruction of a KASETO tape, the cassette is inserted in the control section 4 according to the recognition signal, playback is possible or a check is performed. And if playback is possible, after a response is performed in Japanese from a loudspeaker by the sound signal outputted from the reply signal generating section 5, saying "a tape is played", playback of a tape will begin with it. The case where the cassette is not inserted, "the cassette is not inserted. An audio response is performed saying please insert a cassette."

[0031]

On the other hand, if it utters with "a tape stop" in a play condition, like the case of the aforementioned play directions, a halt is possible at a control section 4, or playback actuation will stop, after outputting the reply signal of "suspending a tape" etc. from the reply signal generating section 5 after a check is performed and carrying out singing of the loudspeaker.

[0032]

Furthermore, if it utters with "tape ejection" to take out a cassette tape, a cassette tape will be ejected after generating the answer tone of "ejecting a tape" by the same principle as the aforementioned case.

[0033]

After engine starting, if the speaker of a front seat 8 is the case where hand free actuation is performed in English, if it utters with "English", the sound signal inputted from Microphones 1a or 1b will input into the language judging section 2 used through the pretreatment circuit 9, and the classification of the language used will be judged to be "English." Only English is confirmed by this judgment signal in the speech recognition section 3 and the reply signal generating section 5, and comparison collating of the dictionary of English registered into the speech recognition section 3 and the sound signal of inputted English is carried out by it. Therefore, the hand free actuation of the loading device can be henceforth carried out in English.

[0034]

Drawing 3 is a flow chart which shows the procedure in the equipment of the example of drawing 2.

Now, in the initial state which started the engine, at step S1, if it utters with "NIHONGO" and voice input is carried out toward the microphones 1a or 1b of drawing 2, a sound signal will output and noise rejection etc. will be performed by the pretreatment circuit 9 of drawing 2 in

step S3 from Microphones 1a or 1b by step S2.

[0035]

And in step S4, the classification of language is judged as the vocabulary in which the sound signal was inputted and inputted into the language judging section 2 by this design being language judging vocabulary, Japanese is set up to a voice recognition unit 3 and the reply signal generating section 5 at step S5, and assignment of the language used is completed.

[0036]

Subsequently, in step S1, if voice input of the operating instructions, such as a "tape play", is carried out, at step S2, a sound signal will output and pretreatment of noise rejection etc. will be performed in step S3 from Microphones 1a or 1b. And it is judged with the vocabulary inputted by the following step S4 not being language judging vocabulary ("NIHONGO" or "English"): That is, since it is the vocabulary of operator command, at step S6, a Japanese dictionary table is chosen by the speech recognition section 3, and the semantics of inputted Japanese is recognized.

[0037]

If it was the case where it had been recognized normally, after playback of a cassette tape will be possible or a check will be performed by the control section 4 in step S8 as a result of the semantic judgment in step S7, in step S9, an audio response is performed saying "A tape is played" and playback actuation of a cassette tape is started in step S10.

[0038]

When judged with the speech recognition in step S6 being impossible at step S7, it sets to step S11, and it is "inputting once again" from the reply signal generating section 5.

A sound signal is outputted, a loudspeaker is driven and the retry of voice actuation is urged.

[0039]

In addition, if an engine is started, it utters with "English" and voice input is carried out at step S1 in an initial state, since it will be judged with "English" in step S4 and English will be set to the speech recognition section 3 and the reply signal generating section 5 in drawing 2, in future voice input, speech recognition and an audio response are performed only in English.

[0040]

In this example, in the initial state which put the engine into operation, it utters with "NIHONGO" or "English", and the future language used is specified first. Therefore, unless an engine is suspended, modification of the language used cannot be performed.

[0041]

On the other hand, if it is made the configuration which excludes judgment step S4 of drawing 3, performs the judgment in the language judging section 2 used of drawing 2, and inputs a language judging signal into the speech recognition section 3 and the reply signal generating section 5 about all the inputted vocabulary. Whenever language is inputted, a language judging can be performed, in the speech recognition section 3, comparison collating can be carried out with the vocabulary in the registration table of the judged language, and an audio response can be performed in the judged language from the reply signal generating section 5. It is not necessary for it not to be necessary to suspend an engine, whenever it changes the language used, and to start again, and to redo a setup of language, and, according to this configuration, operability improves.

[0042]

However, if the vocabulary which is common in Japanese and English is used like operating-instructions words, such as a "tape play" and "tape IJIKUTO", the situation where an audio response is performed in Japanese for the audio response to have been performed in English to have carried out operating instructions in Japanese, or have carried out operating instructions in English conversely will occur, and derangement will be caused. It is necessary to order by the text and, and the language judging section 2 used needs to carry out a language judging like "play a tape" or "eject a tape" using the difference in the functor of not only the description of a word but Japanese, and English etc. without ordering with a word, in order to prevent this.

[0043]

When setting up the language used from input voice artificially like claim 2 by the thought of

claim 1 to being the example which performs a language judging automatically, the above Like drawing 4, if the language used is chosen with a switch etc. in step S0, a setting signal will occur from the setting section 6 of the language used in drawing 1 (2), it will input into the speech recognition section 3 and the reply signal generating section 5, and the future language used will be chosen.

[0044]

Therefore, in step S0, supposing English is set up, when only the operator command in English becomes effective and it uses other language, it is necessary henceforth to reset up in step S0. In this example, once it sets up, as long as there is no modification of the language used, it is not necessary to operate it at all, although it is necessary to operate a language configuration switch manually. You may make it set up a language setup with voice not only a manually-operated switch but as mentioned above.

[0045]

In addition, it is good also as selectable (another side is performed in predetermined language) in the recognition object language of the speech recognition section 3, or the audio response language of the reply signal generating section 5, and you may enable it to set up both language independently separately.

[0046]

not being limited only to this, although the above example is an example which made the two languages of Japanese and English usable — natural — moreover, three or more languages — selection — also suppose that it is usable.

[0047]

The candidate for actuation with voice can perform actuation of various kinds of switches, such as others [ device / in drawing 2 ], for example, a wiper etc., with the voice recognition unit by this design as much as possible. And it is applicable not only to actuation of the device in an automatic in the car one but actuation of other devices.

[0048]

[Effect of the Device]

Since according to claim 1 voice input is only carried out, the language is judged and speech recognition and an audio response are performed in the language, speech recognition and when it can carry out an audio response and the equipment for export and the race of two or more nations use it, it is convenient at two or more languages. Moreover, since it is not necessary to consider as equipment with the export place only for languages, versatility is high and can be realized cheaply.

[0049]

Like claim 2, in a configuration of choosing the language used by switch actuation etc. artificially beforehand, since actuation will not be required unless the language used is changed once it sets up the language used, it is advantageous to little [ the frequency of modification of the language used ] equipment, and since the configuration is simple, it is cheaply realizable.

## DESCRIPTION OF DRAWINGS

---

[Brief Description of the Drawings]

[Drawing 1] It is a block diagram explaining radical Motohara \*\* of the voice recognition unit by this design.

[Drawing 2] It is the block diagram showing the example which carried out the voice recognition unit of claim 1 in the automobile.

[Drawing 3] It is the flow chart which illustrates the procedure in the example equipment of drawing 2.

[Drawing 4] It is the flow chart which illustrates the procedure in the design of claim 2.

[Drawing 5] It is the flow chart which shows the response actuation in the conventional voice recognition unit.

[Description of Notations]

1 Voice Input Section

2 Language Judging Section Used

3 Speech Recognition Section

4 Control Section

5 Reply Signal Generating Section

6 Language Setting Section Used

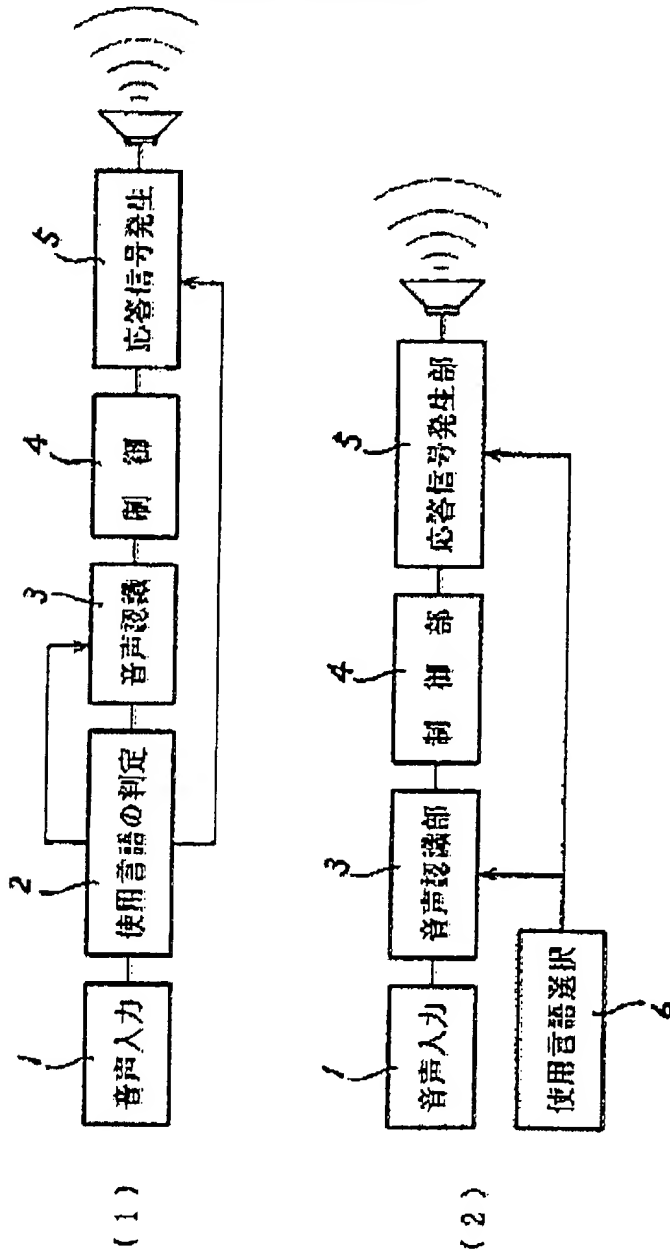
A1-An Various equipment in an automobile



DRAWINGS

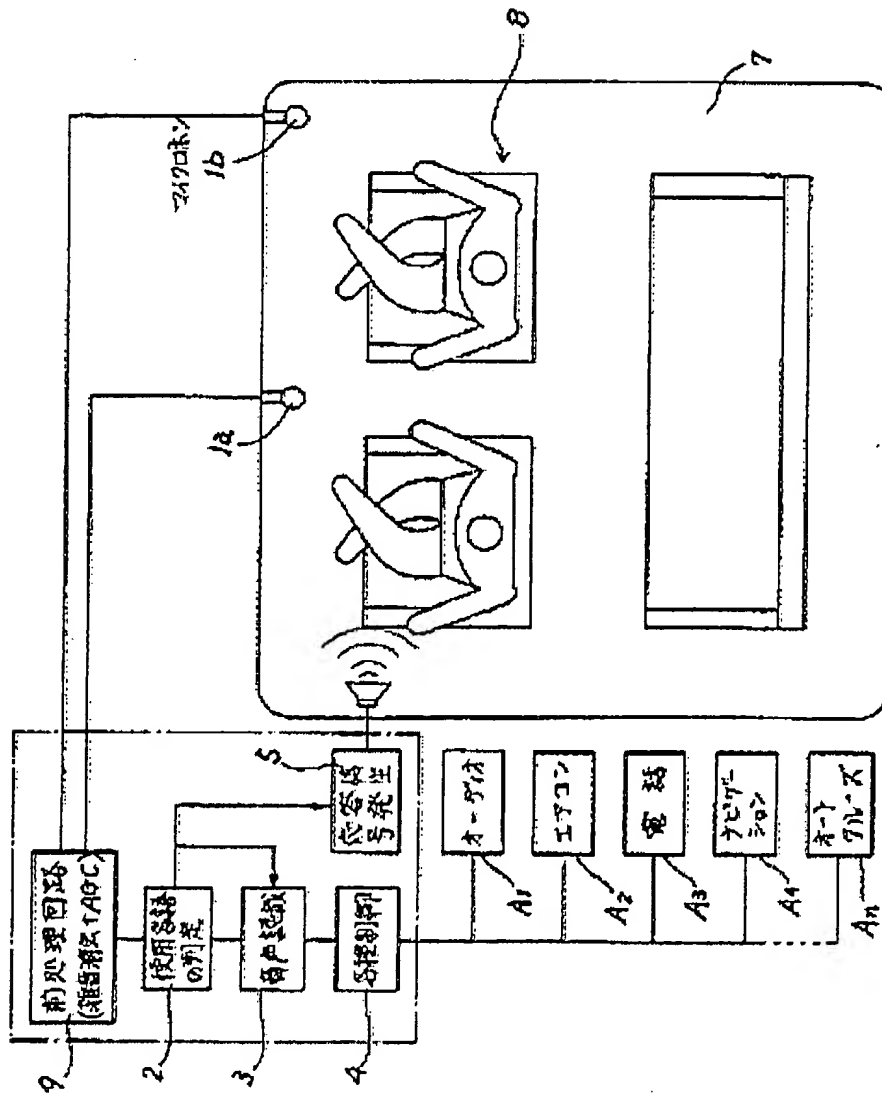
[Drawing 1]

本考案の基本構成



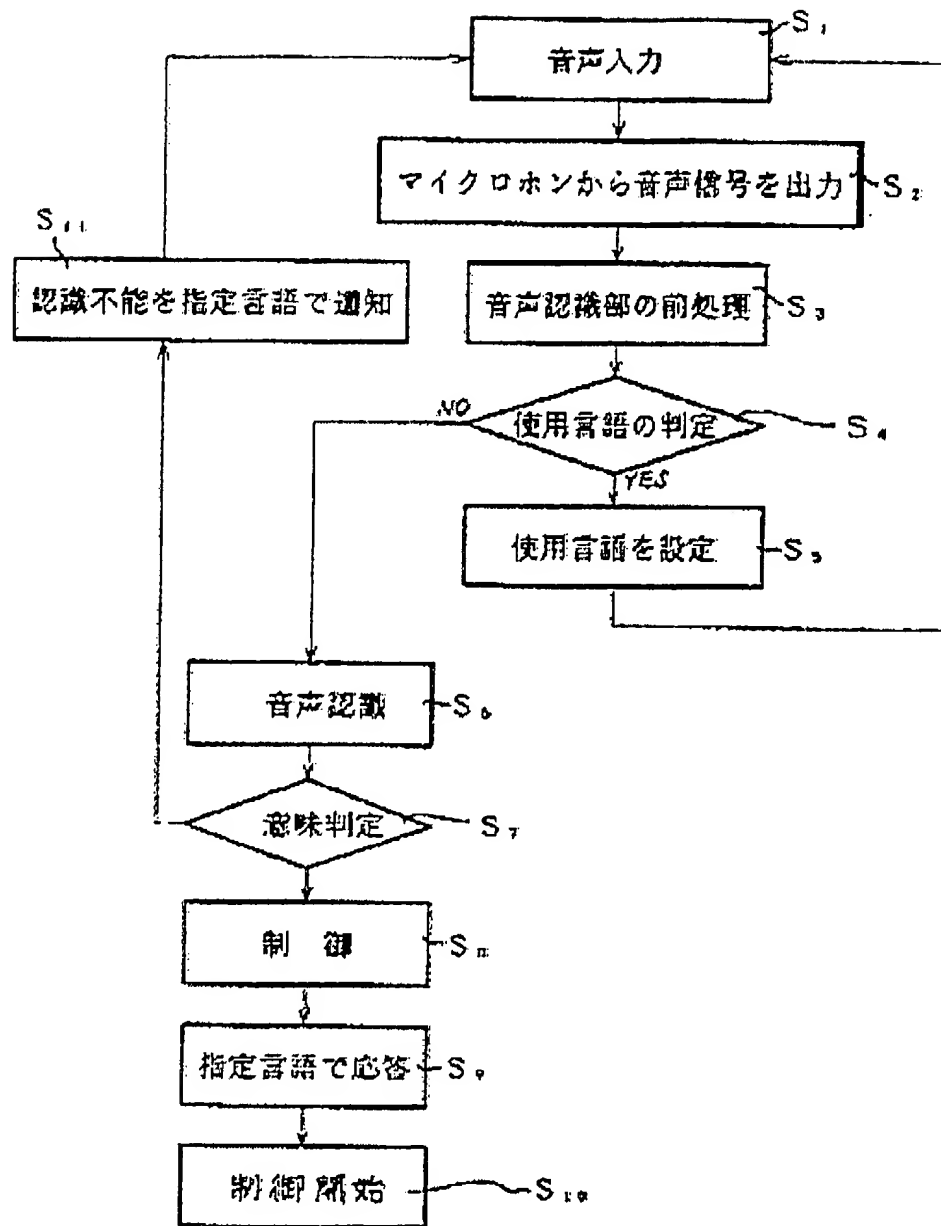
[Drawing 2]

自動車に実施した例



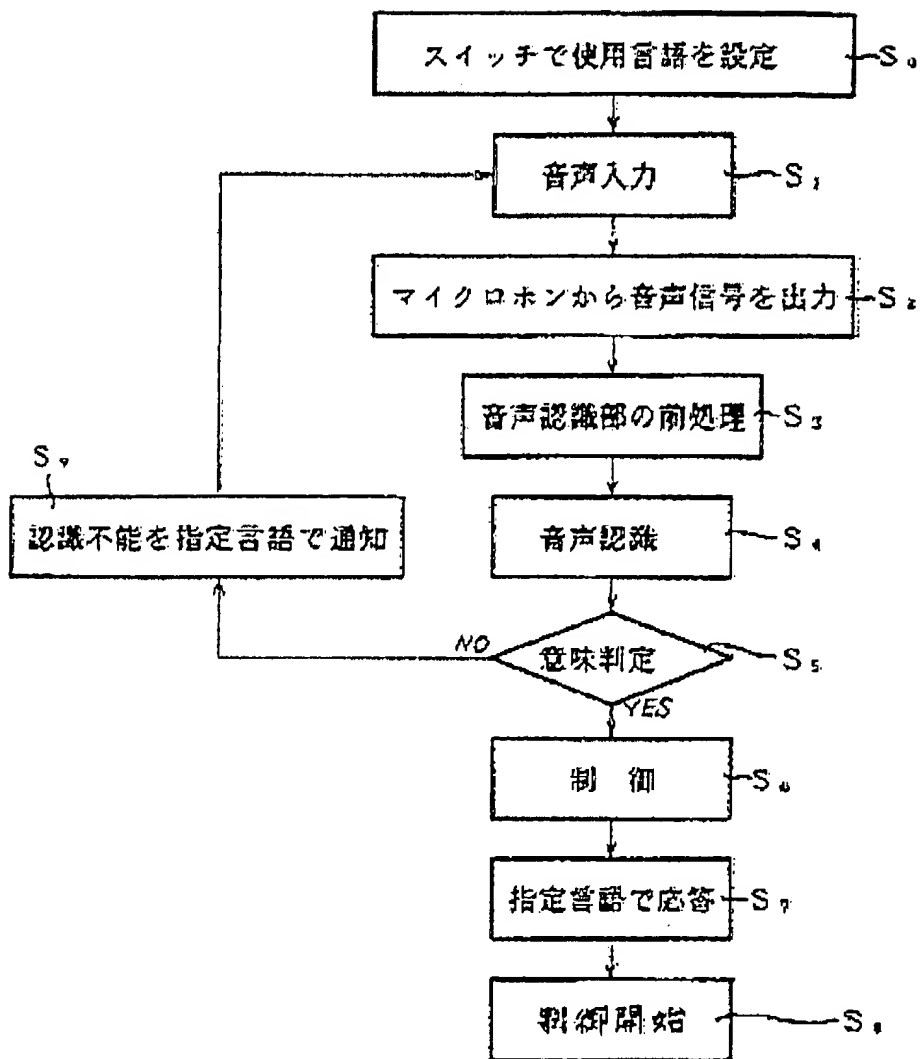
[Drawing 3]

請求項1の実施例



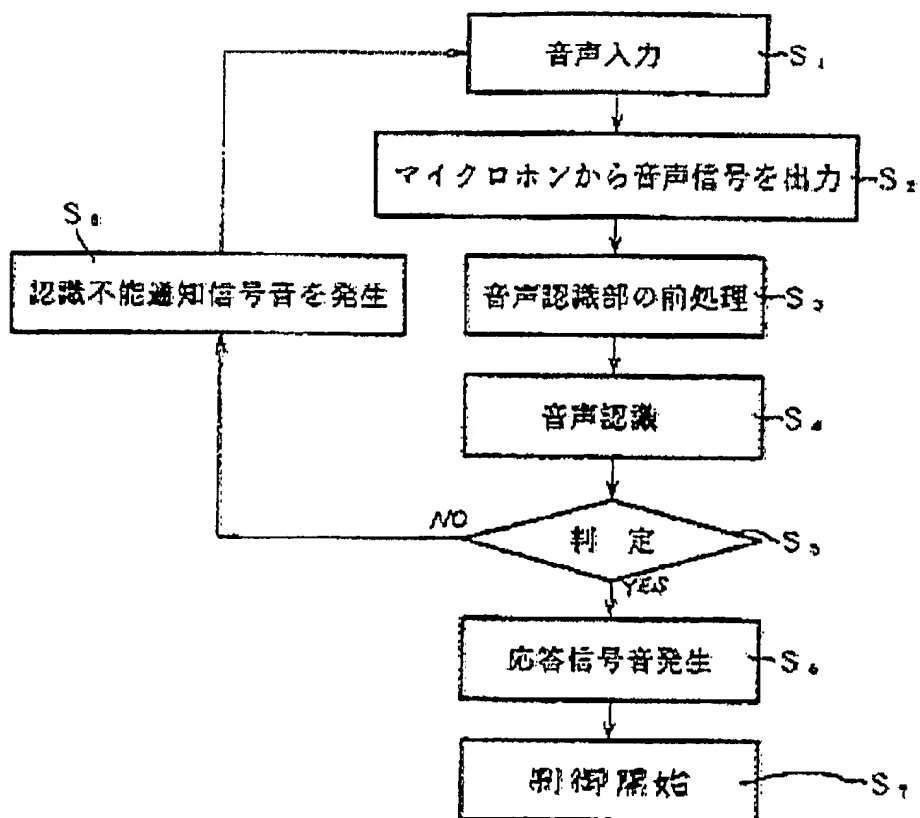
[Drawing 4]

請求項 2 の実施例



[Drawing 5]

## 従来の音声認識装置



(19)日本国特許庁 (J P)

(12) 公開実用新案公報 (U)

(11)実用新案出願公開番号

実開平6-73800

(43)公開日 平成6年(1994)10月18日

(51)Int.Cl. <sup>5</sup>	識別記号	序内整理番号	F I	技術表示箇所
G 1 0 L 3/00	5 6 1 D	9379-5H		
	5 5 1 C	9379-5H		
H 0 4 Q 9/00	3 1 1 Q	7170-5K		

審査請求 未請求 請求項の数 2 O L (全 7 頁)

(21)出願番号 実願平5-15272

(22)出願日 平成5年(1993)3月30日

(71)出願人 000237592

富士通テン株式会社

兵庫県神戸市兵庫区御所通1丁目2番28号

(72)考案者 藤本 博之

兵庫県神戸市兵庫区御所通1丁目2番28号

富士通テン株式会社内

(72)考案者 佐古 和也

兵庫県神戸市兵庫区御所通1丁目2番28号

富士通テン株式会社内

(72)考案者 藤本 昇治

兵庫県神戸市兵庫区御所通1丁目2番28号

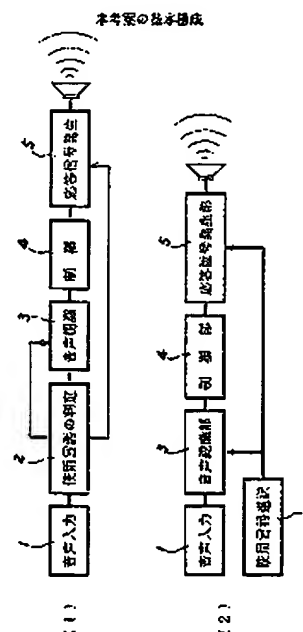
富士通テン株式会社内

(54)【考案の名称】 音声認識装置

(57)【要約】

【目的】話者の発声に対して音声で応答を行なう音声認識装置に関し、予め複数か国語で音声認識および音声応答可能とすることで使い勝手を良くし、かつ汎用性の高い安価な装置を実現することを目的とする。

【構成】少なくとも、音声入力部1と、入力して来た音声信号からその言語を判定する使用言語判定部2と、入力された音声信号の内容を判別する複数か国語対応の音声認識部3と、音声認識結果に従って機器を制御する制御部4と、音声応答のための複数か国語対応の応答信号発生部5とを有し、使用言語判定部2から出力した判定信号が音声認識部3に入力して、認識対象言語が選定され、かつ使用言語判定部2から出力した判定信号が応答信号発生部5に入力して、音声応答言語が選定されるような構成とする。



(2)

実開平6-73800

1

【実用新案登録請求の範囲】

【請求項1】 話者の発声した音声の信号が入力すると、該音声信号に対応するトークバック信号を出力し、音声で応答する音声認識装置において、

少なくとも、音声入力部(1)と、入力した音声信号からその言語を判定する使用言語判定部(2)と、入力された音声信号の内容を判別する音声認識部(3)と、音声認識結果に従って機器を制御する制御部(4)と、音声応答のための応答信号発生部(5)とを有し、音声認識部(3)は、入力した音声信号から複数カ国語を認識でき、また応答信号発生部(5)は、複数カ国語による音声応答が可能となっており、

使用言語判定部(2)から出力した判定信号が音声認識部(3)に入力して、認識対象言語が選定され、かつ使用言語判定部(2)から出力した判定信号が応答信号発生部(5)に入力して、音声応答言語が選定されるように構成されていることを特徴とする音声認識装置。

【請求項2】 話者の発声した音声の信号が入力すると、該音声信号に対応するトークバック信号を出力し、音声で応答する音声認識装置において、

少なくとも、音声入力部(1)と、入力された音声信号の内容を判別する音声認識部(3)と、音声認識結果に従って機器を制御する制御部(4)と、音声応答のための応答信号発生部(5)と、使用言語の設定部(6)とを有し、音声認識部(3)は、入力した音声信号から複数カ国語

2

を認識でき、また応答信号発生部(5)は、複数カ国語による音声応答が可能となっており、

前記音声認識部(3)および前記応答信号発生部(5)の少なくとも一方は、使用言語設定部(6)から出力された選択信号に応じて、認識対象言語あるいは音声応答言語が選択されるように構成されていることを特徴とする音声認識装置。

【図面の簡単な説明】

【図1】本発明による音声認識装置の基本原理を説明するブロック図である。

【図2】請求項1の音声認識装置を自動車に実施した例を示すブロック図である。

【図3】図2の実施例装置における処理手順を例示するフローチャートである。

【図4】請求項2の発明における処理手順を例示するフローチャートである。

【図5】従来の音声認識装置における応答動作を示すフローチャートである。

【符号の説明】

- 1 音声入力部
- 2 使用言語判定部
- 3 音声認識部
- 4 制御部
- 5 応答信号発生部
- 6 使用言語設定部

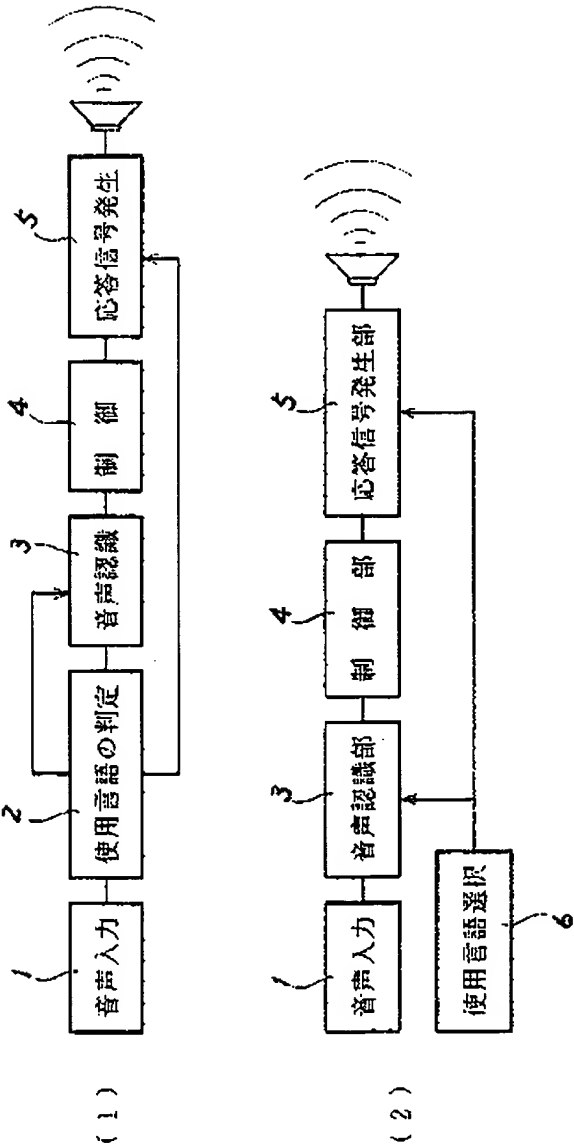
A1～An 自動車における各種装置

(3)

実開平6-73800

【図1】

本考案の基本構成



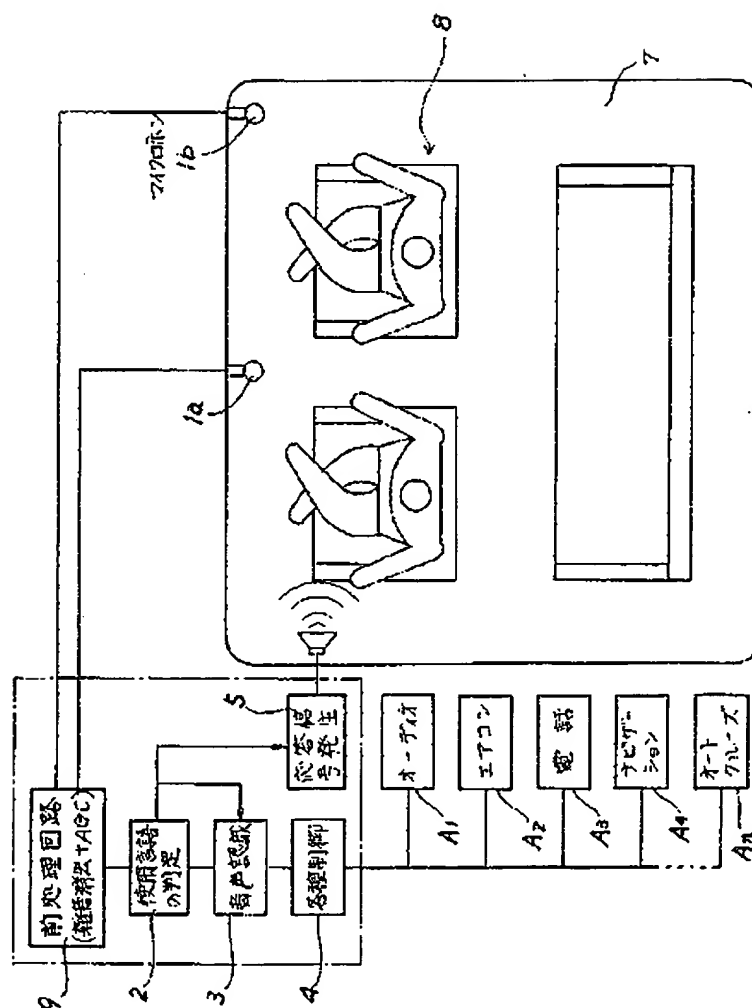


(4)

実開平6-73800

【図2】

自動車に実施した例

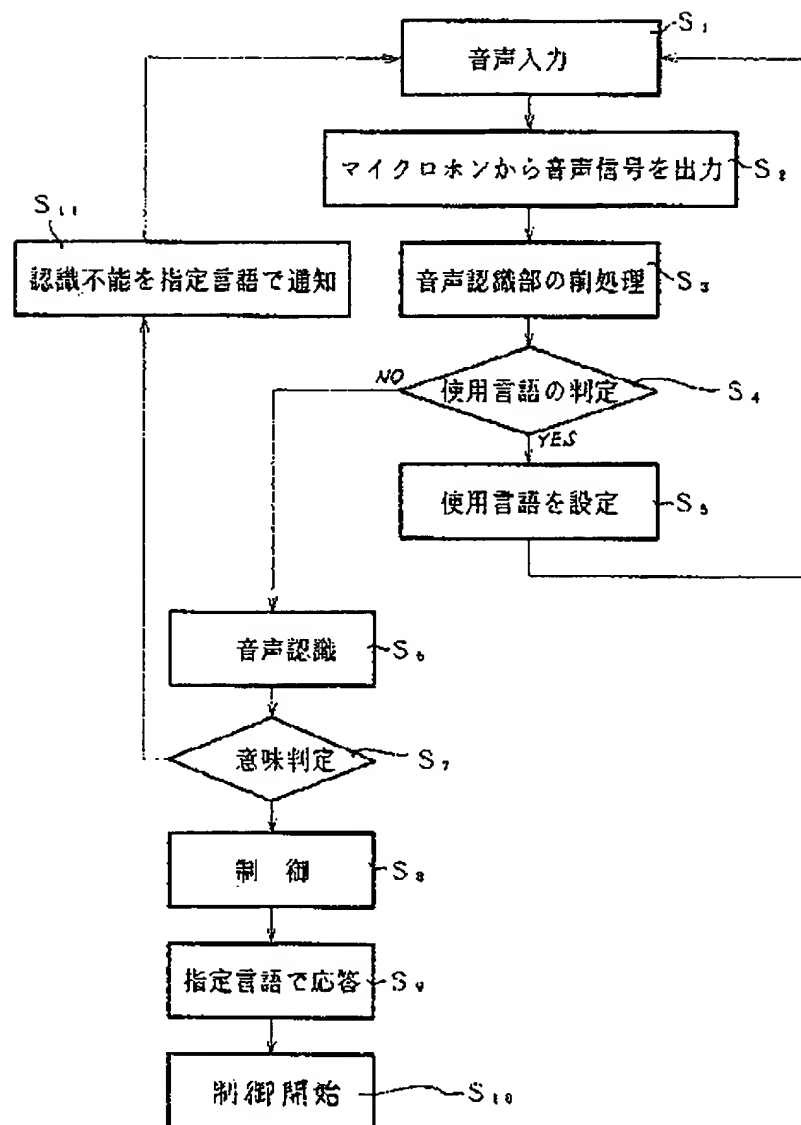


(5)

実開平6-73800

【図3】

請求項1の実施例

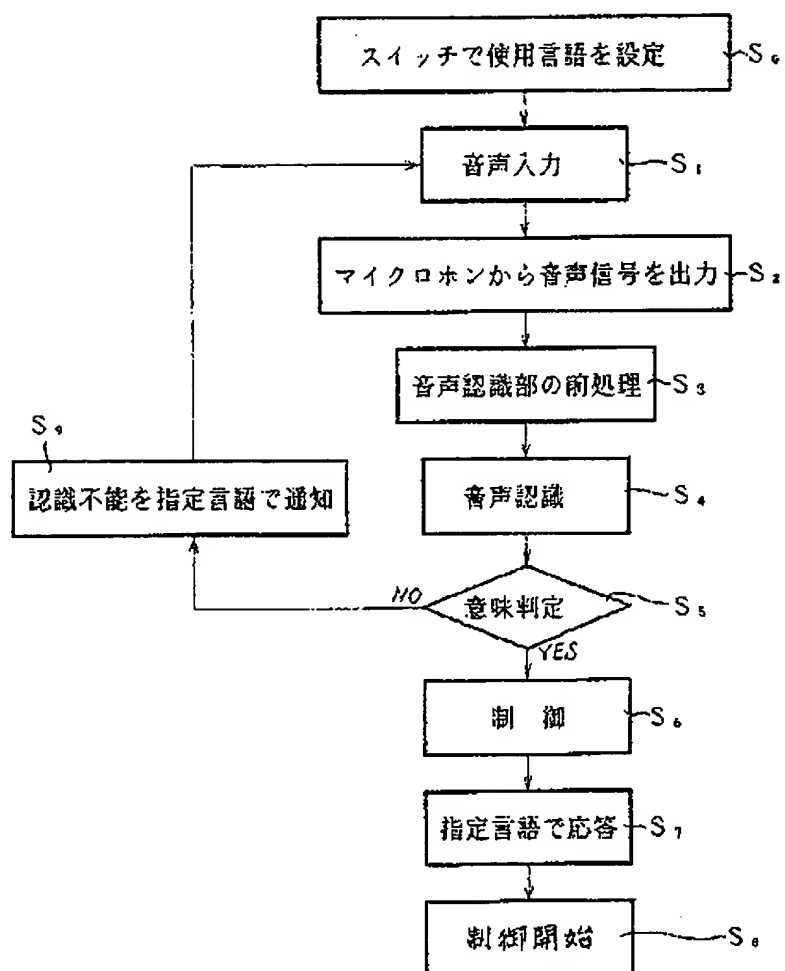


(5)

実開平6-73800

【図4】

請求項2の実施例

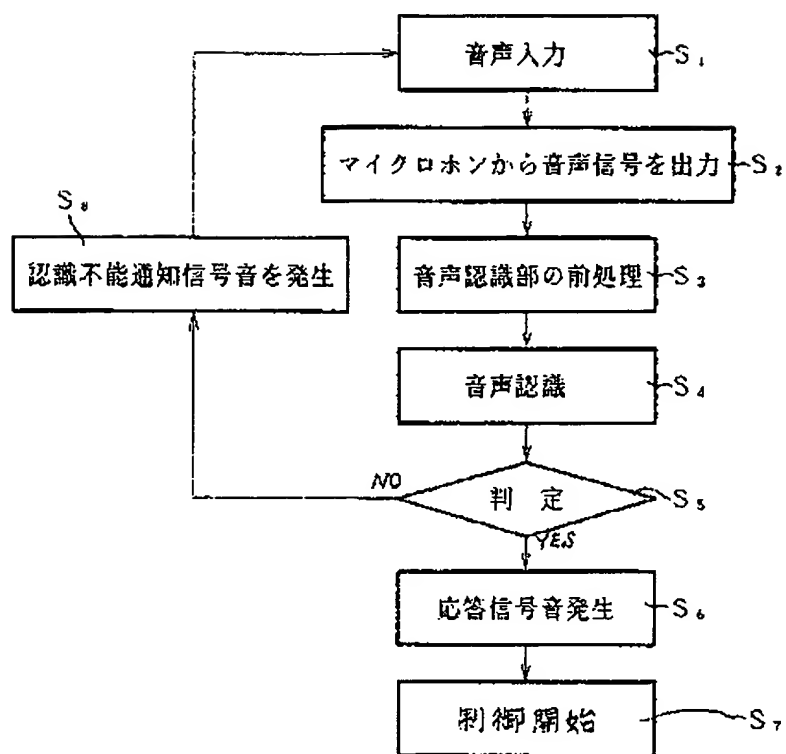


(7)

実開平6-73800

【図5】

従来の音声認識装置



(8)

実開平6-73800

## 【考案の詳細な説明】

## 【0001】

## 【産業上の利用分野】

本考案は、話者の発声に対して音声で応答を行なう音声認識装置に関する。自動車などの運転者が運転中に、エアコンディショナやオーディオ装置、カーナビゲーション、自動車電話、オートクルーズなどの機器を手動操作すると、安全運転に支障を来し、事故を引き起こすおそれがある。そこで、これらの機器の操作を音声で指示し、また機器側では、指示された操作に対して音声で応答して、話者が確認できるようにすることが、操作性向上のために望ましい。

## 【0002】

## 【従来技術】

自動車用に限らず、音声認識装置は、話者の音声を認識し、さらに認識した内容に従って機器を操作するようになっている。そして、結果を話者に対して、機器側から信号音を発して応答する。

## 【0003】

図5は従来の音声認識装置における応答動作を示すフローチャートである。オーディオ機器を操作する場合であれば、まず、ステップS1において、例えば「プレイ」と発声して音声入力し、再生動作を指示すると、ステップS2において、マイクロホンを通じて音声信号が入力され、ステップS3で音声認識に適するように前処理が行なわれる。

## 【0004】

次いで、ステップS4に示すように、音声認識部において、音声認識が行なわれ、ステップS5において認識結果の判定が行なわれる。「プレイ」という言葉を認識できた場合は、ステップS6において、認識できたことをブザー音などで応答して話者に知らせた後、ステップS7において、プレイ動作を開始する。

## 【0005】

これに対し、話者の言葉を認識できない場合は、ステップS8において、その旨を知らせる信号音を出力して、話者に対し、再度音声入力するよう促す。

## 【0006】

(9)

実開平6-73800

## 【考案が解決しようとする課題】

しかしながら、このようなブザー音などによる応答では、音声認識できたか、できないかの二通り程度しか表現できない。自動車などの場合は、多数の機器が装備されており、多数のスイッチを操作する必要があるので、スイッチ操作によって行なわれる動作の内容も多種多様である。したがって、応答もブザー音ではなく、操作指示の内容に応じて具体的に音声で応答できるのが望ましい。

## 【0007】

例えば、話者が「プレイ」と発声した場合、その指示音声を確認し、しかもプレイ可能であれば、例えば「カセットテープをプレイします」のように、音声で応答し、音声認識できない場合は、「もう一度お話し下さい」のように、音声応答（トークバック）できると、対話式に、より正確に機器操作を行なうことができ、操作性が向上する。

## 【0008】

また、エアコンディショナの場合は、温度や風量などの状態を、オートクルーズの場合は、現在の速度などの状態を、音声応答で話者に知らせることができると、操作がより簡便となる。

## 【0009】

ところで、音声認識は通常、日本語で行なわれるが、輸出車の場合は、相手国語で音声認識可能にしておく必要がある。また、本考案のように音声で応答する場合も、輸出用の車両については、相手国語で応答する必要がある。

## 【0010】

しかしながら、音声認識やトークバックを、輸出車に関してのみ、英語などの相手国語で行なえるようにするには、英語を音声認識可能とし、しかもトークバックも英語で行なえる認識装置を別に開発する必要がある、コスト高となるという問題がある。

## 【0011】

一方、国際化が進展すると、レンタルカーなどを想定したとき、同一車両を日本人が運転することもあれば外国人が運転することもある。このような場合、日本語のみ又は英語のみしか音声認識やトークバックできないとなると、所期の目

(10)

実開平6-73800

的を達成できず、不便である。

#### 【0012】

本考案の技術的課題は、このような問題に着目し、予め複数カ国語で音声認識および音声応答可能とすることで使い勝手を良くし、かつ汎用性の高い安価な装置を実現することにある。

#### 【0013】

##### 【課題を解決するための手段】

図1は本考案による音声認識装置の基本構成を説明するブロック図である。請求項1の音声認識装置は、話者の発声した音声の信号が入力すると、該音声信号に対応するトークバック信号を出力し、音声で応答する音声認識装置において、図1(1)のように、少なくとも、音声入力部1と、入力して来た音声信号からその言語を判定する使用言語判定部2と、入力された音声信号の内容を判別する音声認識部3と、音声認識結果に従って機器を制御する制御部4と、音声応答のための応答信号発生部5とを有している。

#### 【0014】

前記の音声認識部3は、入力した音声信号から複数カ国語を認識でき、また応答信号発生部5は、複数カ国語で音声応答できる。そして、使用言語判定部2から出力された判定信号が音声認識部3に入力して、認識対象言語が選定され、かつ使用言語判定部2から出力された判定信号が応答信号発生部5に入力して、音声応答言語が選定されるように構成されている。

#### 【0015】

請求項2は、話者の発声した音声の信号が入力すると、該音声信号に対応するトークバック信号を出力し、音声で応答する音声認識装置において、図1(2)のように、少なくとも、音声入力部1と、入力された音声信号の内容を判別する音声認識部3と、音声認識結果に従って機器を制御する制御部4と、音声応答のための応答信号発生部5と、使用言語の設定部6とを有している。

#### 【0016】

音声認識部3は、入力した音声信号から複数カ国語を認識でき、また応答信号発生部5は、複数カ国語による音声応答が可能となっている。そして、前記音声

(11)

実開平6-73800

認識部3および前記応答信号発生部5の少なくとも一方は、使用言語設定部6から出力された選択信号に応じて、認識対象言語あるいは音声応答言語が選択されるように構成されている。

【0017】

【作用】

本考案の音声認識部3は、それぞれの言語ごとに音声認識機能を有していて、使用言語判定部2からの判定信号によって、判定された言語の音声認識機能が作動する構成になっている。したがって、入力した音声信号から複数カ国語を認識できる。

【0018】

また、応答信号発生部5は、それぞれの言語ごとに、操作命令に応じた音声応答機能を有しているため、使用言語判定部2において、入力して来た音声信号からその言語が自動的に判定されて、判定信号が音声認識部3と応答信号発生部5に入力し、使用言語が指定されると、音声認識部3では指定された言語のみを選択して音声認識し、応答信号発生部5では、指定された言語を選択して音声応答が行なわれる。

【0019】

このように、音声入力部1から音声が入力するだけで、その言語が判定され、その言語で音声認識や音声応答が行なわれるので、複数カ国語で音声認識および音声応答でき、輸出用の装置や複数カ国の人種が使用する場合に便利である。また、輸出先の国語専用の装置とする必要がなく、汎用性が高いので、安価に実現できる。

【0020】

なお、使用言語判定部2においては、初期状態で入力してきた最初の音声信号で言語の種別を判定して、以後はその言語のみを有効とする手法と、音声信号が入力してくるたびにその言語の種別を判定し、判定信号を音声認識部3および応答信号発生部5に入力して、使用言語を指定する手法とがあるが、本考案は両方を含むものとする。

【0021】



(12)

実開平6-73800

請求項2における音声認識部3も、入力した音声信号から複数カ国語を認識でき、また応答信号発生部5は、複数カ国語で音声応答が可能となっており、使用言語設定部6から出力した設定信号が音声認識部3に入力して、認識対象言語が選択され、あるいは使用言語設定部6から出力した設定信号が応答信号発生部5に入力して、音声応答言語が選択される構成になっている。

【0022】

したがって、予め人為的に使用言語を選択する操作が必要ではあるが、複数カ国語に対応できるので、請求項1の場合と同様な効果を奏するほか、一旦使用言語を設定すれば、使用言語を変更しない限り、操作を行なう必要がないので、使用言語の変更の頻度が少ない装置に有利であり、また構成が簡素となるため、より安価に実現できる。

【0023】

【実施例】

次に本考案による音声認識装置が實際上どのように具体化されるかを実施例で説明する。図2は請求項1の音声認識装置を自動車に実施した例を示すブロック図である。自動車7のフロントシート8の前方には、運転席側および助手席側の双方に、音声入力手段として、マイクロホン1a、1bが配設されている。

【0024】

そして、マイクロホン1a、1bで検出された音声信号は、雑音除去などを行なう前処理回路9を経て、使用言語の判定部2に入力され、フロントシート8における話者の話した言語の種別が判別される。使用言語判定部2において、話者の使用言語が判別されると、その判別信号が音声認識部3と応答信号発生部5に入力され、判定された言語のみが有効となる。

【0025】

この自動車には、装備として、オーディオ装置A1、エアコンディショナA2、自動車電話A3、ナビゲーション装置A4およびオートクルーズA<sub>n</sub>などが搭載されている。

【0026】

この実施例は、初期状態で入力してきた最初の音声信号で言語の種別を判定し

(13)

実開平6-73800

て、以後はその言語のみを有効とする構成になっているものとする。いま、フロントシート8の話者が、日本語で機器の操作を行なう場合であれば、エンジンを始動させて初期状態において、まず「ニホンゴ」と発声すると、マイクロホン1aまたは1bから入力した音声信号は、前処理回路9を経て、使用言語判定部2に入力し、使用言語の種別が「日本語」とであると判定される。

【0027】

そして、この判定信号によって、音声認識部3および応答信号発生部5では、日本語のみが有効とされ、音声認識部3に登録されている日本語の単語ないし文章と、入力してきた単語ないし文章とが比較照合される。

【0028】

例えば、オーディオ装置A1のカセットテープを再生する場合であれば、話者がまず「テープ・プレイ」と発声すると、マイクロホン1aまたは1bから入力した音声信号は、前処理回路9および使用言語判定部2を経て、音声認識部3に入力し、音声認識部3に登録されている日本語の辞書と比較照合される。

【0029】

英語など、他の国語の辞書も登録されているが、他の国語が登録されているテーブルは選択されず、使用言語判定部2から入力した判定信号によって、日本語が登録されているテーブルのみが選択され、予め登録されている用語と比較照合される。

【0030】

その結果、カセットテープの再生命令であることが認識され、認識信号にしたがって制御部4において、カセットが挿入されていて再生が可能か確認が行なわれる。そして、再生が可能であれば、応答信号発生部5から出力する音声信号によって、例えば「テープをプレイします」と、スピーカから日本語で応答が行なわれた後、テープの再生が開始する。カセットが挿入されていない場合は、「カセットが挿入されていません。カセットを挿入して下さい」と、音声応答が行なわれる。

【0031】

一方、プレイ状態において、例えば「テープ・ストップ」と発声すると、前記

(14)

実開平6-73800

のプレイ指示の場合と同様にして、制御部4で停止が可能か確認が行なわれた後、応答信号発生部5から「テープを停止します」等の応答信号を出力してスピーカを鳴動させた後に、再生動作が停止する。

## 【0032】

さらに、カセットテープを取り出したい場合は、「テープ・イジェクト」と発声すれば、前記の場合と同じ原理で「テープをイジェクトします」等の応答音を発生した後に、カセットテープがイジェクトされる。

## 【0033】

フロントシート8の話者が、英語でハンドフリー操作を行なう場合であれば、エンジン始動後に、例えば「イングリッシュ」と発声すると、マイクロホン1aまたは1bから入力した音声信号が前処理回路9を経て、使用言語判定部2に入力し、使用言語の種別が「英語」と判定される。この判定信号により、音声認識部3および応答信号発生部5では、英語のみが有効とされ、音声認識部3に登録されている英語の辞書と、入力してきた英語の音声信号とが比較照合される。したがって、以後は英語によって、搭載機器をハンドフリー操作できる。

## 【0034】

図3は、図2の実施例の装置における処理手順を示すフローチャートである。いま、エンジンを始動させた初期状態において、ステップS1で、図2のマイクロホン1aまたは1bに向かって、「ニホンゴ」と発声し音声入力すると、ステップS2でマイクロホン1aまたは1bから音声信号が出力し、ステップS3において図2の前処理回路9により雑音除去などが行なわれる。

## 【0035】

そして、ステップS4において、本考案による言語判定部2に音声信号が入力され、入力された用語が言語判定用語であると、言語の種別が判定されて、ステップS5で音声認識装置3および応答信号発生部5に対して、日本語が設定され、使用言語の指定が完了する。

## 【0036】

ついで、ステップS1において、例えば「テープ・プレイ」などの操作命令が音声入力されると、ステップS2でマイクロホン1aまたは1bから音声信号が出力

(15)

実開平6-73800

し、ステップS3において雑音除去などの前処理が行なわれる。そして、次のステップS4で、入力された用語は言語判定用語（“ニホンゴ”または“イングリッシュ”）ではないと判定される。すなわち、操作指令の用語であるので、ステップS6で、音声認識部3により、日本語の辞書テーブルが選択されて、入力した日本語の意味が認識される。

【0037】

ステップS7における意味判定の結果、正常に認識された場合であれば、ステップS8において制御部4によりカセットテープの再生が可能か確認が行なわれた後、ステップS9において、「テープを再生します」と音声応答が行なわれ、ステップS10において、カセットテープの再生動作が開始される。

【0038】

ステップS6における音声認識が不可能であると、ステップS7で判定された場合は、ステップS11において、応答信号発生部5から「もう一度入力して下さい」との音声信号を出力してスピーカを駆動し、音声操作のリトライを促す。

【0039】

なお、エンジンを始動させて初期状態において、ステップS1で「イングリッシュ」と発声し音声入力すると、ステップS4において「英語」と判定されて、図2における音声認識部3および応答信号発生部5には、英語が設定されるので、以後の音声入力では、英語のみで音声認識および音声応答が行なわれる。

【0040】

この実施例では、エンジンを始動した初期状態において、「ニホンゴ」または「イングリッシュ」と発声して、最初に、以後の使用言語を指定するようになっている。したがって、エンジンを停止しない限り、使用言語の変更はできない。

【0041】

これに対し、図3の判定ステップS4を省き、入力された用語のすべてについて、図2の使用言語判定部2における判定を行ない、音声認識部3および応答信号発生部5に言語判定信号を入力する構成にすれば、言語が入力される度に言語判定を行ない、音声認識部3において、判定された言語の登録テーブル中の用語と比較照合し、かつ応答信号発生部5からは判定された言語で音声応答を行なう

(15)

実開平6-73800

ことができる。この構成によると、使用言語を変更するたびに、エンジンを停止し再度始動して言語の設定をやり直す必要はなく、操作性が向上する。

## 【0042】

しかしながら、「テープ・プレイ」や「テープ・イジェクト」などの操作命令語のように、日本語と英語で共通している用語を用いると、日本語で操作命令したのに英語で音声応答が行なわれるとか、逆に英語で操作命令したのに日本語で音声応答が行なわれる、といった事態が発生し、混乱を来す。これを防止するには、単語で命令しないで、「テープをプレイしなさい」や「テープをイジェクトしなさい」などのように、文章で命令する必要がある、また使用言語判定部2は、単語の特徴のみでなく、日本語と英語の構文の違いなどを利用して言語判定する必要がある。

## 【0043】

以上は、請求項1の思想によって、入力音声から自動的に言語判定を行なう例であるのに対し、請求項2のように人為的に使用言語を設定する場合は、図4のように、ステップS0において、スイッチなどで使用言語を選択すると、図1(2)における使用言語の設定部6から設定信号が発生して、音声認識部3および応答信号発生部5に入力し、以後の使用言語が選択される。

## 【0044】

したがって、ステップS0において、例えば英語が設定されたとすると、以後は英語による操作指令のみが有効となり、他の言語を用いる場合は、ステップS0において設定しなおす必要がある。この実施例では、言語設定スイッチを手動操作する必要があるが、一旦設定しておく、と、使用言語の変更がない限り何ら操作を行なう必要がない。言語設定は、手動操作スイッチに限らず、前記のように例えば音声で設定するようにしてもよい。

## 【0045】

なお、音声認識部3の認識対象言語と応答信号発生部5の音声応答言語のいずれか一方のみを選択可能（他方は所定言語で行なう）としてもよく、また両者の言語を別個に独立して設定できるようにしてもよい。

## 【0046】

(17)

実開平6-73800

以上の実施例は、日本語と英語の二カ国語を使用可能とした例であるが、これのみに限定されないことはもちろんであり、また三カ国語以上を選択使用可能とすることもできる。

【0047】

音声による操作対象は、図2における機器のほか、例えばワイパーなどのような各種のスイッチ類の操作を、可能な限り本考案による音声認識装置で行なうことができる。しかも、自動車内における機器の操作に限らず、他の機器の操作にも適用できる。

【0048】

【考案の効果】

請求項1によれば、音声入力するだけで、その言語が判定され、その言語で音声認識や音声応答が行なわれるので、複数カ国語で音声認識および音声応答でき、輸出用の装置や複数カ国の人種が使用する場合に便利である。また、輸出先の国語専用の装置とする必要がないので、汎用性が高く、安価に実現できる。

【0049】

請求項2のように、予め人為的にスイッチ操作などで使用言語を選択する構成の場合は、一旦使用言語を設定すれば、使用言語を変更しない限り、操作を要しないので、使用言語の変更の頻度が少ない装置に有利であり、構成が簡素なためより安価に実現できる。